

Adaptive Modeling and Anti-error Correction Algorithm for Abnormal Behavior of Power Load Based on Reinforcement Learning

Liang Xiao^{1,a}, Guanghu Xu^{1,b}, Shijin Xu^{1,c}, Quan Yuan^{1,d} and Xiaobo Zhang^{2,e*}

¹China Southern Power Dispatching and Control Center, Guangzhou 510700, Guangdong, China

²China Southern Power Grid Artificial Intelligence Technology Co., Ltd, Guangzhou 510700, Guangdong, China

In regard to the high uncertainty and error correction risk problems in the identification of abnormal behavior of power load, this paper proposes an adaptive modeling and anti-error correction algorithm based on reinforcement learning. In the adaptive modeling stage, a Markov decision process with random perturbations is constructed; multi-scale state representation is integrated; and the load fluctuation characteristics are dynamically captured. In the anti-error correction stage, a hierarchical safety constraint action space and an interpretable reward function are designed, and the attention-driven strategy network is combined to focus on key features. A dynamic safety barrier module based on load fluctuation entropy is applied to achieve risk perception and real-time interception of high-risk correction. Experiments show that the retention rate of the proposed model reaches 0.87, and the correction success rate is 92% under a noise intensity of 20db, which is up to 22% higher than the baseline model.

Keywords: reinforcement learning; power load anomaly identification; adaptive modeling; anti-error correction; multi-scale state representation

1. INTRODUCTION

The precise identification and dynamic correction of abnormal behavior of power loads are of great significance in smart grids and power system security [1–2]. As the scale of power systems continues to expand and the complexity of operation increases, load forecasting and anomaly detection have become key considerations to ensure the stable operation of power grids and improve energy efficiency. However, existing research generally has two limitations: on the one hand, traditional methods mostly use static thresholds or fixed parameter models, which are difficult to dynamically adapt

to the time evolution characteristics of sudden anomalies in load sequences; on the other hand, correction strategies generally lack hierarchical safety constraint mechanisms and are prone to triggering chain risks due to misjudgment under difficult working conditions. Load data is often mixed with random fluctuations and nonlinear trends. It is difficult to adapt to the rapid changes and uncertainties in actual working conditions by relying solely on traditional static models. Hence, it is essential to design intelligent algorithms with adaptive capabilities and safety constraints to achieve timely discovery and robust correction of abnormal load behaviors. The identification and correction of power load anomaly is subject to high uncertainty, noise interference, and potential error-correction risks. Load changes are affected by multiple factors such as seasonality, meteorological conditions, and

*Corresponding author. ^aEmail: xiaoliang@csg.cn, ^bxguangh@126.com, ^cmike_xushijin@foxmail.com, ^dyuanquan529@126.com, ^exiaobozh_aitech@163.com.

user behavior patterns, resulting in multi-scale non-stationary characteristics in load sequences [3–4]. The parameters of traditional static models are fixed, and it is impossible to adjust feature weights in real-time according to the changes in the entropy value of load fluctuations. Its correction method usually relies on deterministic rules and fails to establish a dynamic safety barrier to prevent high-risk operations. In the correction process, if there is a lack of effective safety constraints and dynamic feedback mechanisms, it is easy to over-correct or fail to identify real anomalies in time, which hurts the safety and operating efficiency of the system. These problems have led to the need to accurately model the load state under complex working conditions and achieve a balance between sensitivity and safety during anomaly correction, which has become the main difficulty in current research.

In response to these challenges, researchers in recent years have applied deep neural networks and time series modeling methods to load forecasting and anomaly detection [5–6], extracting local time series features through convolutional neural networks, capturing long-term dependencies using long short-term memory networks, and combining graph neural networks to process spatial correlations. At the same time, some methods try to apply adversarial training or data enhancement to improve the robustness of the model to abnormal patterns, and some methods implement anomaly detection and correction based on supervised learning or unsupervised clustering [7–9]. However, most methods rely on static model parameters and cannot dynamically adapt to changes in load states. At the same time, correction strategies often lack interpretability and safety constraints, and there are still problems such as error correction risks and insufficient model generalization.

This paper innovatively combines Markov decision process modeling with random perturbations and multi-scale state representation to achieve the dynamic capture of complex load fluctuations; a hierarchical safety constraint action space is designed to avoid the risk of misoperation, and an interpretable reward function is constructed to precisely guide strategy optimization. The application of an attention-driven strategy network focuses on key abnormal features to improve decision transparency, and is supplemented by a dynamic safety barrier module to intercept high-risk corrections in real-time. Experimental verification shows that the framework significantly surpasses the baseline model in terms of recognition precision, response speed, and operational safety (F1 up to 0.89/drift adaptation 28 seconds/correction success rate 92%), providing an end-to-end solution for high-uncertainty power scenarios that combines adaptability, interpretability, and security.

2. RELATED WORK

In recent years, the research on power load forecasting and anomaly detection has gradually become a hot topic in the field of smart grids. Shirzadi et al. developed and compared different regional power load forecasting models by using multiple methods such as support vector machines, random forests, nonlinear autoregressive exogenous neural

networks, and long short-term memory networks [10]. The results showed that deep learning methods were superior to other methods in the accuracy of forecasting load demand, but their prediction stability still faced certain challenges when facing missing and incomplete data. In contrast, Dong et al. proposed a short-term power load forecasting method based on support vector machines, based on K-Means clustering. By considering temperature and holiday factors and using K-Means to group seasonal load data, the prediction precision and operation efficiency were improved. However, it relied on a fixed load grouping method and lacked flexibility when facing dynamic and frequently changing load fluctuations [11]. To solve this problem, Oprea et al. proposed a hybrid method based on unsupervised machine learning, combining a spectral residual convolutional neural network and a Martingale-based anomaly detection model to successfully detect abnormal power consumption behavior, demonstrating the potential of unsupervised learning in power data anomaly detection [12]. However, it relied on significant abnormal patterns and might perform poorly in the face of more complex and dynamic data environments. To further explore online learning methods, Ahmed et al. proposed a real-time training method based on deep learning, PMUNET (Phasor Measurement Unit Network), which detected, located, and classified anomalies on streaming PMU data through two variants: dynamic data change-driven learning and continuity-driven learning [13]. This method improved the responsiveness to dynamic data through a real-time training mechanism, but under high load conditions, the pressure of real-time computing and data processing was still a major challenge. As the research deepened, Takiddin et al. proposed a method for detecting power theft network attacks based on strong autoencoders and long short-term memory sequence-to-sequence structures [14]. Simulation experiments found that although the autoencoder with a seq2seq (Sequence to Sequence) structure could better capture the complex patterns of power consumption data, the stability of the model might still be affected when encountering noisy or missing data.

In regard to smart grids, Udo et al. explored the application of machine learning technology in load forecasting, demand response management, and energy efficiency optimization, focusing on how machine learning can improve load forecasting precision, achieve dynamic demand adjustment, and enhance grid stability [15]. Although case studies have shown that this approach achieved good results in a variety of application scenarios, how to deal with the real-time processing of large-scale, multi-source data remains a difficult problem to solve. With the increasing severity of smart grid security issues, Presekal et al. proposed a hybrid deep learning method based on a graph convolutional long short-term memory network and a deep convolutional network for online detection of network attacks in power systems, and combined with software-defined networks, communication throughput anomaly detection, and attack graph models, successfully achieved high-precision positioning of attack locations [16]. However, although this method improves the precision of attack positioning, its complex network structure and high computational overhead still face major challenges in its actual deployment in large-scale power systems. At

the same time, to solve the real-time detection problem, Jahangir et al. proposed a 2DR-CNN (2-D convolutional neural network classifier and reconstruction decoder) method based on a multi-output network, which used the generator bus frequency and phase angle data monitored by the phasor measurement unit to detect and locate dynamic load change attacks with high resolution [17]. Simulation results showed that this method was superior to other machine learning methods in detecting and locating D-LAAs (dynamic load-altering attacks) and could effectively deal with PMU (phasor measurement unit) measurement noise and missing data. Even though existing research methods have made significant progress in power load forecasting, anomaly detection, and security protection, they are still unable to deal efficiently with load fluctuations, anomaly detection, and data missing in complex and dynamic power systems.

In the field of power load anomaly identification, many studies have explored the application of reinforcement learning and deep learning to enhance the security and stability of the power grid. Mohamed & Kundur proposed a method based on deep reinforcement learning to identify potential vulnerabilities in power grid load frequency control and simulate multiple attack strategies to help formulate defense strategies and optimize attack detection methods [18]. This method uses deep reinforcement learning to identify potential network vulnerabilities and simulate attack strategies, providing an effective defense idea for the power grid. In further research, Lei et al. proposed a method based on reinforcement learning to help prevent load redistribution attacks by identifying and reinforcing key branches of the power grid, and enhanced the defense capabilities of the power grid in terms of operational economy and safety through cross-update search strategies [19]. By applying a cross-update search strategy, this approach showed more comprehensive and effective protection performance in preventing LRA (load redistribution attacks), showing the application potential of reinforcement learning in dynamic load management. Shereen et al. proposed to use the proximal strategy optimization algorithm to learn effective attack strategies by modeling undetectable attacks as a multi-objective partially observable Markov decision process, and proved that these attacks could cause the power system frequency to exceed the critical limit without being discovered by traditional detection methods [20]. Gou et al. proposed a method for detecting abnormal power consumption behavior of low-voltage power users based on the K-GBDT (K-Gradient Boosting Decision Tree) hybrid model, which combined the global modeling capability of the GBDT (Gradient Boosting Decision Tree) algorithm and the local classification advantage of the KNN (K-Nearest Neighbor) algorithm [21]. Aljohani et al. proposed an intrusion detection and mitigation system based on deep learning neural networks and long short-term memory algorithms, which could effectively detect, classify, and locate intrusion behaviors in smart grids, and diagnose and predict contaminated smart electronic equipment [22]. By combining DLNN (deep learning neural networks) and LSTM (long short-term memory), this study further enhanced the anomaly detection and error correction capabilities of the power system, providing a guarantee for the robustness of the power grid. Cooper et al. reviewed the basic meth-

ods and latest technologies for anomaly detection through power system state estimation, discussed traditional bad data detection methods and state estimation anomaly detection methods based on data-driven rapid change detection and artificial intelligence, and proposed research directions for future smart grids [23]. Alam et al. revealed the potential of machine learning models to improve network security and operational efficiency in smart grids, and pointed out challenges such as data privacy, scalability, and integration with traditional systems [24]. They proposed solutions using technologies such as federated learning and blockchain, and recommended further research to improve the robustness and efficiency of the model.

Based on these challenges, the further development of the error correction mechanism needs to find a balance between the interpretability of the model and the security of data to achieve more efficient and secure power system management. Although previous research has provided a variety of reinforcement learning and deep learning methods for power load anomaly identification and defense, how to balance computational efficiency with the real-time and precision of error correction in specific applications remains an issue to be resolved.

3. ALGORITHM DESIGN

3.1 Dynamic Modeling of Environment

To cope with the high uncertainty and random fluctuations of abnormal behavior of a power load, this study models the dynamic changes of power load by constructing a Markov decision process of the power system with random fluctuations [25–26]. This process utilizes the historical data and current state of power load, and dynamically adjusts the model to cope with the real-time changes of load fluctuations, ensuring that the system can self-adapt and respond during anomaly detection and correction. In specific implementation, the environmental state $s_{env,t}$ is defined as the time series information of each load node in the power system and its corresponding characteristic values, such as load change rate, power factor, frequency fluctuation, etc. The state space combines the time domain characteristics of load data with the frequency domain characteristics to form a multi-dimensional representation, which can comprehensively indicate the operating status of the power system. In addition, the environment also includes contextual information related to external factors, such as climate change, equipment operating status, etc., which are dynamically integrated into the state model to enhance the system's responsiveness and prediction accuracy. For each moment t , the environmental state $s_{env,t}$ can be expressed as:

$$s_{env,t} = (\Delta P_t, \Delta Q_t, \frac{\partial P}{\partial t}, \frac{\partial Q}{\partial t}, \alpha_t, \beta_t) \quad (1)$$

In Formula (1), ΔP_t and ΔQ_t represent the time domain changes of power; $\frac{\partial P}{\partial t}$ and $\frac{\partial Q}{\partial t}$ represent the rates of change of power; α_t and β_t represent the frequency domain characteristics of the load. To further improve the accuracy

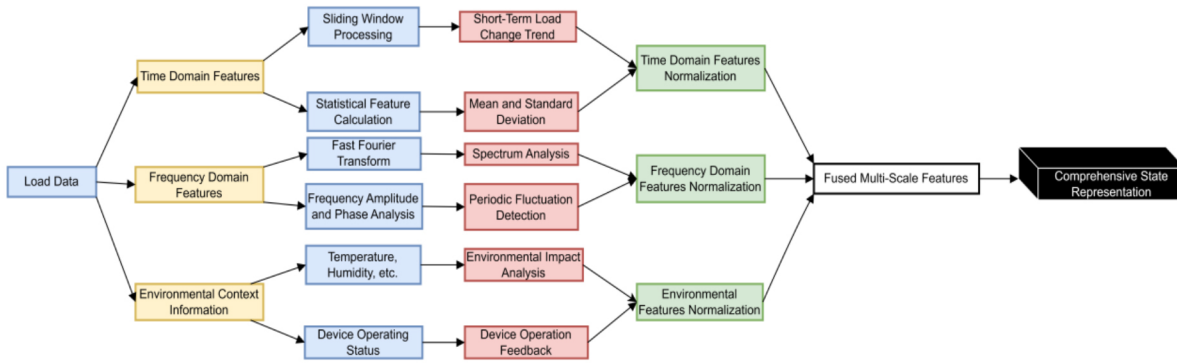


Figure 1 Multi-scale state representation flow chart.

of anomaly detection, this paper adopts a random fluctuation modeling method based on a Markov process. The random disturbance term ξ_t is applied to the system dynamic model to simulate the unpredictable factors in load change. The modeling assumption of the disturbance term ξ_t is a normal distribution with a mean of 0 and a variance of σ_t^2 , that is:

$$\xi_t \sim N(0, \sigma_t^2) \quad (2)$$

By applying random disturbances that conform to the normal distribution, the sudden abnormal fluctuations in the load data are effectively simulated. This explicit modeling of the uncertainty of abnormal behavior is the key basis for building an adaptive model to dynamically capture anomalies. At the same time, the modeling of the disturbance term takes into account the statistical characteristics of load fluctuations, which can effectively capture the nonlinear changes and abnormal fluctuations of power loads, thereby providing richer contextual information for the decision-making process. In the Markov decision process, the system state $s_{sys,t}$ is defined as the processing state of the power load, covering operations such as observation, marking, correction, and manual review. The action selected at each moment is determined by the system state $s_{sys,t}$, and the current strategy $\pi(a_t|s_{sys,t})$. The strategy $\pi(a_t|s_{sys,t})$ is a conditional probability distribution that describes the probability of taking different actions under a given system state:

$$\pi(a_t|s_{sys,t}) = P(a_t|s_{sys,t}) \quad (3)$$

The agent learns by maximizing the cumulative reward. The cumulative reward R_t is defined as the weighted sum of the current instant reward and future rewards:

$$R_t = r_t + \gamma E[R_{t+1}|s_{sys,t}, a_t] \quad (4)$$

In Formula (4), r_t is the immediate reward at the current moment; γ is the discount factor; $E[R_{t+1}|s_{sys,t}, a_t]$ represents the expected future reward after selecting action a_t under the system state $s_{sys,t}$. To ensure the safety and effectiveness of the correction process, the action selection function considers the impact of each action on the stability of the system and limits the possible risk of error correction. In the action space design, by setting safety constraints, the action selection is limited to an appropriate range to avoid overcorrection or application of error correction.

3.2 Multi-Scale State Representation

During the identification of abnormal behavior and the prevention of error correction of power loads, load data usually presents multi-scale characteristics, including time domain changes, frequency domain characteristics, and correlation with environmental factors. To accurately capture this information and improve the model's recognition precision, this study designs a multi-scale state representation method to fuse different levels of information of load data to form a more comprehensive state description [27–28].

The time domain characteristics process the instantaneous changes of load data, including the rate of change and fluctuation amplitude of the load. These characteristics reflect the dynamic changes of the load within a short period of time and are the basis for identifying rapid fluctuations and sudden anomalies. By means of the sliding window method, the load change trend in the short term is obtained in realtime, and its statistical characteristics are calculated to provide a time series basis for subsequent abnormal judgment. The frequency domain characteristics start from the frequency components of the load data and use the fast Fourier transform to analyze the spectrum of the load signal [29–30]. Frequency domain analysis can effectively reveal the periodic changes and extra-periodic interference in load fluctuations. Especially when the power system is subject to periodic load changes or equipment failures, frequency domain features can significantly improve the accuracy of anomaly detection. This part of the features indicates the deep structure of load fluctuations through the combination of frequency, amplitude, and phase difference. To strengthen the adaptability to environmental changes, environmental context information is also incorporated into the state representation. These factors may have a direct or indirect impact on load changes. Through the fusion of environmental context data, the model can consider changes in the external environment when detecting load anomalies and avoid misidentification caused by environmental factors. All these time domain, frequency domain, and context features are normalized during state representation to ensure that features at different scales have the same dimensions and weights, thereby avoiding conflicts or imbalances between features. The relevant process is shown in Figure 1:

Table 1 Safety constraint action space design table.

Action Level	Action Definition	Execution Conditions	Safety Constraints
Observation	Collect real-time data, monitor state	System status is normal, data is available	No constraints
Marking	Mark the load behavior as normal or abnormal	Based on historical data and feature analysis	No correction operation, pure marking
Correction	Adjust load, perform correction	Triggered when an anomaly is marked	Correction magnitude is limited by the load change rate
Manual Review	The expert confirms if the correction is valid	Triggered when an anomaly is complex or uncertain	Must be reviewed by experts before performing the correction

3.3 Safety Constraint Action Space

It is vital to ensure safety during power load anomaly identification and correction. This study designs an action space with safety constraints to prevent the algorithm from applying unnecessary risks or error corrections when performing anomaly correction. The design of the action space is subdivided through hierarchical action sets, including four operation levels: observation, marking, correction, and manual review. Each level contains different action constraints and execution conditions. The core of this hierarchical design is the building of a progressive safety barrier: low-risk operations are executed first to provide a decision-making basis and safety buffer for high-risk operations; high-risk operations are subject to strict physical constraints; the highest risk decisions apply manual intervention as the ultimate guarantee, thereby systematically reducing the risk of error correction.

The entire action space design is centered on safety constraints to ensure that the risks of misoperation and system crash can be effectively avoided at each stage of action execution. The implementation of the correction operation is carried out strictly in accordance with the safety specifications of the power system to avoid the negative impact of over-correction on system stability. The design of this hierarchical action set enables the algorithm to automatically select the most appropriate operation according to different abnormality levels and load conditions, thereby maximizing the stability and reliability of the system while ensuring the correction effect. Its structure is shown in Table 1:

Table 1 shows the design of the safety constraint action space in the power load anomaly correction process. Each action level (observation, marking, correction, and manual review) has clear definitions, execution conditions, and corresponding safety constraints. Observation and marking operations do not involve system correction and are used mainly for data collection and anomaly marking; correction operations are limited by the load change rate to ensure that they do not cause system instability. Manual review provides the final verification when the anomaly is complex to avoid incorrect correction. This design ensures the safety and stability of the load correction process.

3.4 Interpretable Reward Function

This study proposes an interpretable reward function that combines detection precision, a false positive penalty, and

safety risk to ensure that the algorithm maintains system stability and safety while effectively correcting. The core goal of the reward function is to maximize the accuracy of load anomaly detection. To quantify the detection precision, the precision factor is applied, which is defined as:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

In Formula (5), TP represents the number of times the real anomaly is correctly detected, and FP represents the number of times the normal load is misjudged as an anomaly. To enhance the precision of the system, the precision factor is incorporated into the reward function, and a positive reward value is assigned:

$$R_{precision} = \lambda_1 \cdot Precision \quad (6)$$

In Formula (6), λ_1 is the weight coefficient of the precision factor. Adjusting this coefficient can balance the influence of other objectives. False positive penalty is the key to ensuring system reliability. This study designs a false positive penalty term:

$$R_{penalty} = -\lambda_2 \cdot \frac{FP}{TP + FN} \quad (7)$$

In Formula (7), FN represents the number of times the actual abnormality is not detected, and λ_2 is the penalty coefficient, which controls the penalty intensity of false positives.

During the correction operation, the safety of the system must also be guaranteed. This study considers the safety risks and quantifies the potential risks caused by the correction based on the load fluctuation entropy. Fluctuation entropy reflects the randomness and uncertainty of load changes and is defined as:

$$H = -\sum_{i=1}^n p(x_i) \cdot \log(p(x_i)) \quad (8)$$

In Formula (8), $p(x_i)$ is the probability distribution of load changes, and n is the size of the load state space. The design of the safety risk term is as follows:

$$R_{risk} = -\lambda_3 \cdot H \quad (9)$$

In Formula (9), λ_3 is the weight coefficient of safety risk. Finally, the reward function of comprehensive precision, false positive penalty, and safety risk can be expressed as:

$$R_{total} = R_{precision} + R_{penalty} + R_{risk} \quad (10)$$

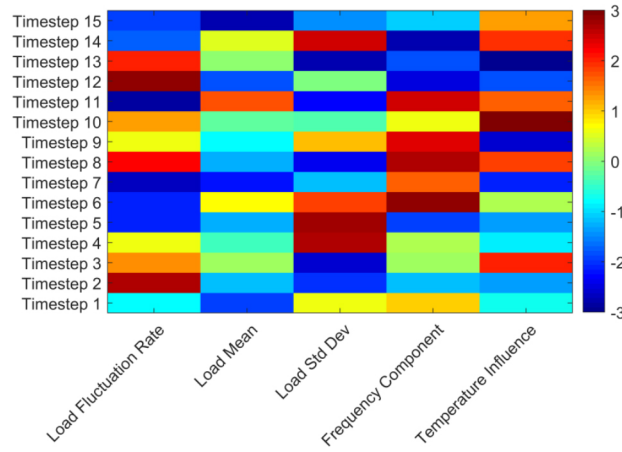


Figure 2 Generating a heat map of the decision basis.

3.5 Attention-Driven Policy Network

To improve the decision-making ability and transparency of load anomaly identification and correction, this study designs a policy network based on the attention mechanism, which combines a convolutional neural network and a self-attention mechanism. This attention mechanism not only improves the model's ability to understand the input data, but also enhances the transparency of the decision-making process. By generating a heat map of the decision basis, the system can show those time steps or features that have a greater impact on the final decision result. This interpretability feature helps analyze and optimize model performance, and also provides visual feedback to power system engineers to help them understand the decision basis of the model and further tune it. During the training process, the policy network guides parameter updates according to the reward function, so that the network not only pays attention to the precision of load anomaly detection, but also optimizes the correction strategy. Through the attention-driven policy network, the algorithm can adjust its decision strategy in realtime to ensure the best operation under different load anomaly types and system states, thereby ensuring the stability and security of the power system. The results are shown in Figure 2:

Figure 2 shows the attention weights generated when modeling the abnormal behavior of power load based on reinforcement learning. The ordinate represents the time step (Timestep), and the abscissa represents different power load characteristics (such as load fluctuation rate, load mean, load standard deviation, frequency component, and temperature influence). Some time steps (such as the load data in step 5) fluctuate greatly, indicating the dynamic characteristics of load changes. The reasons may be related to seasonal changes, equipment load, and other factors. Time steps with a high load fluctuation rate may correspond to a sudden increase in equipment load, causing abnormal fluctuations.

3.6 Safety Barrier Module

When correcting abnormal power loads, ensuring safety is crucial. To avoid system instability or wrong decision-

making caused by the correction process, this paper proposes a dynamic safety barrier module based on load fluctuation entropy.

The module evaluates the stability of the current system by calculating the load fluctuation entropy (LFE). Fluctuation entropy is an important indicator of the complexity and uncertainty of load data, which can indicate the disorder of the system load. For the load data sequence $L(t) = \{l_1(t), l_2(t), \dots, l_n(t)\}$ at a certain moment, the entropy value is defined as:

$$H(L) = - \sum_{i=1}^n p_i \log(p_i) \quad (11)$$

In Formula (11), p_i represents the probability distribution of the load value $l_i(t)$ in historical data.

The core of the safety barrier module is to set a dynamic threshold $T_{safe}(t)$ according to the fluctuation entropy $H(L)$. When the load fluctuation entropy exceeds the threshold, the system considers that the load abnormal correction operation at the current moment has a high risk and the safety barrier needs to be triggered. The threshold $T_{safe}(t)$ is adjusted according to the changes in real-time load data and the state estimation of the system. The formula for the adjustment is:

$$T_{safe}(t) = \alpha H_{\max} + \beta \cdot \Delta L(t) \quad (12)$$

In Formula (12), $H_{\max}$ is the maximum value of the historical load fluctuation entropy; $\Delta L(t)$ is the fluctuation difference between the current load and the historical load; α and β are adjustment coefficients. Through this dynamic adjustment mechanism, the safety barrier module can optimize the safety threshold in real-time according to the changes in load fluctuation. When the safety threshold $T_{safe}(t)$ is triggered, the module automatically activates the high-risk warning mechanism and suppresses the current correction operation. The correction decision of the model enters the manual review stage or turns to a more conservative observation strategy. This strategy effectively prevents the system from making wrong corrections in complex or high-risk situations. The module also iteratively optimizes the safety threshold through the feedback mechanism and the execution

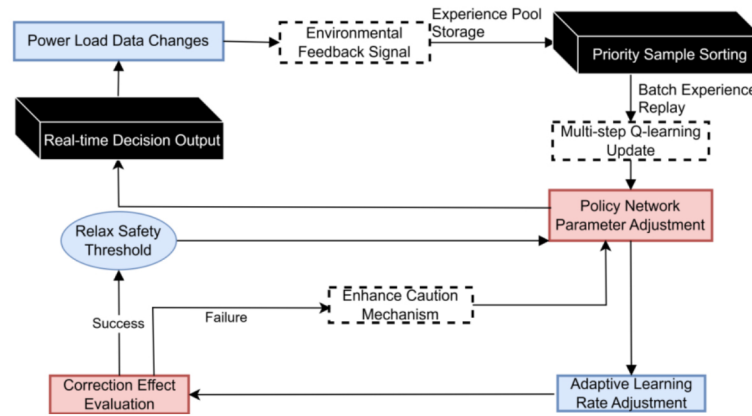


Figure 3 Strategy update mechanism.

results of the correction behavior. If a correction leads to a significant improvement in system stability, the safety barrier threshold can be appropriately relaxed; conversely, if an error correction occurs, the caution of the correction behavior is enhanced by increasing the safety threshold.

3.7 Online Strategy Update Mechanism

To ensure that the algorithm can adapt to the dynamic changes of the environment during the identification and correction of power load anomalies, especially the drift of load abnormal behavior patterns, this study designs an online strategy-update mechanism. This mechanism adjusts the network parameters of the reinforcement learning model in realtime so that it can self-optimize according to the new load data and system environment, thereby improving the model's adaptive capture ability and correction accuracy for dynamic anomalies.

The core idea of the online strategy update mechanism is to use environmental feedback signals to fine-tune the current strategy. Whenever the load data changes, the system re-evaluates the effect of the existing strategy and adjusts the model's behavior strategy based on the evaluation results. This dynamic adjustment capability is crucial so as to effectively respond to evolving abnormal behaviors, ensuring that the model does not fail due to environmental changes, thereby maintaining high-precision identification and low-risk correction. This process is implemented through a multi-step Q-learning algorithm, which combines the execution results of the current strategy with the long-term rewards of the environment to update the strategy parameters to ensure that the decision-making process is always optimized.

In terms of specific implementation, the model adopts a learning method based on experience replay. Whenever the system observes a new load anomaly, it stores the behavior, state, and reward in the experience pool. Through batch replay, the reinforcement learning model can gradually learn the optimal strategy to adapt to environmental changes in multiple updates. In this process, the sample update adopts a priority experience replay mechanism, which gives priority to

those abnormal data samples with greater impact for training, thereby accelerating the learning process. The online strategy update mechanism also combines an adaptive learning rate adjustment strategy. During periods of drastic load data changes, the system automatically increases the learning rate to quickly respond to system mutations; in stable periods, the learning rate automatically decreases to ensure that the model does not over-adjust, and maintains long-term stability. This adjustment strategy effectively prevents the model from over fitting or ignoring long-term returns.

The update of the model not only depends on the immediate feedback of load anomaly detection, but also includes the feedback evaluation of the system after the correction operation. When the correction result deviates from expectations, the system uses this information to adjust the model's decision-making process through back-propagation. This mechanism enables the algorithm to learn from the correction results, continuously optimize its strategy, and significantly reduce the risk of error correction in the future. Through this dynamic adjustment mechanism, the algorithm can self-correct under continuous environmental interaction and improve accuracy and robustness under different load fluctuations. Its strategy update mechanism is shown in Figure 3:

4. EXPERIMENT AND VERIFICATION

4.1 Experimental Design

To verify the effectiveness of the proposed adaptive modeling and anti-error correction algorithm for abnormal behavior of power load based on reinforcement learning, this study constructs a mixed dataset containing five types of anomalies to simulate various complex abnormal situations in the power load system. The dataset contains normal load data and five typical anomalies: sudden load peak, long-term load fluctuation, load drift, sudden equipment failure, and load fluctuation caused by external interference factors.

The normal load samples in the dataset are generated by collecting and processing historical power load data

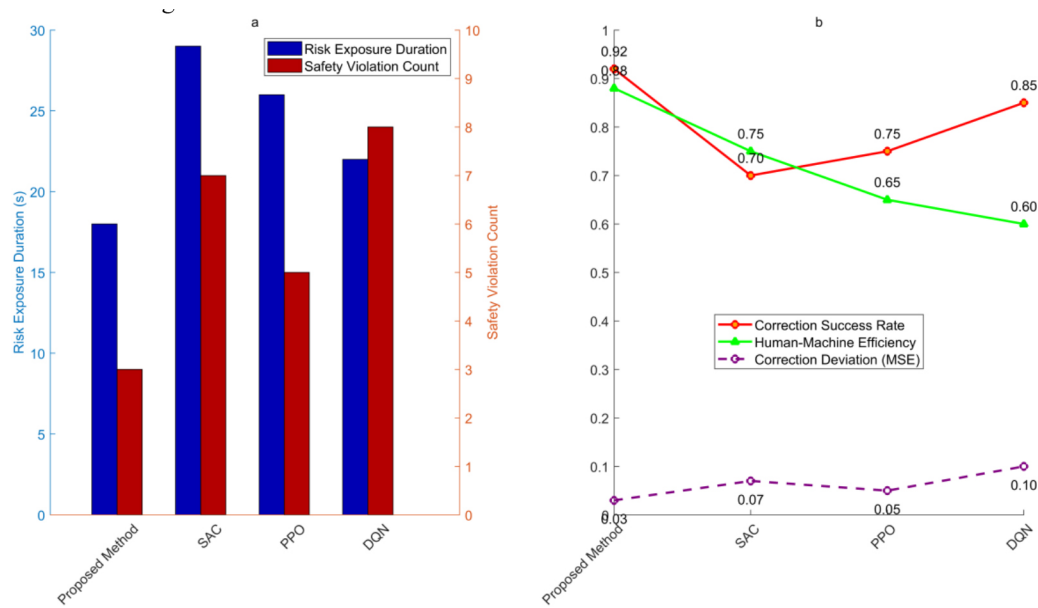


Figure 4 Comparison of core indicators of safety performance. (a) Comparison of safety and risk exposure time. (b) Analysis of correction performance and human-machine collaboration.

and combining the load change characteristics in the actual operating environment. To increase the diversity and authenticity of the experiment, abnormal load data is generated by applying artificially simulated power load interference, equipment failure, and other factors. Each type of anomaly is modeled through a specific change pattern: for example, load surge is simulated by peak superposition; equipment failure is simulated by load fluctuation after a sudden power outage; external interference is generated by random disturbance mode. All data are preprocessed in time series and divided into a training set and a test set. The training set is used for model learning and optimization, and the test set is used to evaluate the detection precision and robustness of the model. The dataset is divided into several time periods, and the load state of each time period is divided by multi-scale state representation to better capture the diversity of abnormal fluctuations. At each time point, the state characteristics of the power load data are converted into the input of the reinforcement learning model. After the dataset has been constructed, the Markov decision process is used for modeling. By modeling the dynamic evolution of abnormal behavior, the system can adaptively learn through randomly fluctuating environmental feedback signals. In each round of learning, the model evaluates its corresponding action strategy according to the current load state and then adjusts the strategy through reward signal feedback. During the training process, through the online learning mechanism, the model can continuously adjust its behavior strategy to handle different types of abnormal situations. To verify the real-time impact and effectiveness of the algorithm, the system simulates anomaly detection and correction under different load fluctuation conditions. Through experiments on real-time response and detection delay of multiple anomalies, the adaptability of the model in the face of new types of load anomalies is evaluated. Considering the dynamic characteristics of the power system, the dataset also contains different load pattern drifts. The experiment further verifies

the robustness and stability of the algorithm in the face of long-term changing load data through the detection of concept drift and performance recovery evaluation. To comprehensively evaluate the performance of the algorithm, the experimental setting compares the model in this paper with three mainstream reinforcement learning baseline models: DQN (Deep Q-learning), SAC (Soft Actor-Critic), and PPO (Proximal Policy Optimization).

4.2 Correction Safety and Efficiency Evaluation

Ensuring the safety of the abnormal correction process is a core requirement for power system operation and maintenance. The experiment described in this section was conducted to quantitatively evaluate the safety performance and risk control ability of the algorithm when performing load correction operations. The experiment designs multi-dimensional safety indicators and verifies the effectiveness of the safety constraint mechanism in this paper by comparing the performance of different algorithms on these key safety and efficiency indicators. The results are shown in Figure 4:

Figure 4 shows the performance of four different algorithms in safety performance and correction tasks. In Figure (a), the number of safety violations of the proposed model is 3, and the risk exposure time is 18 seconds; the number of safety violations of the DQN algorithm is 8, and the risk exposure time is 22 seconds, indicating that the proposed model performs better in these two aspects. Figure 4(b) shows the comparison of correction success rate, collaborative efficiency, and correction amplitude deviation (mean squared error, MSE). The correction success rate of the proposed model is 0.92; the collaborative efficiency is 0.88; the MSE is 0.03, which are significantly better than the results obtained by the other three algorithms, indicating that

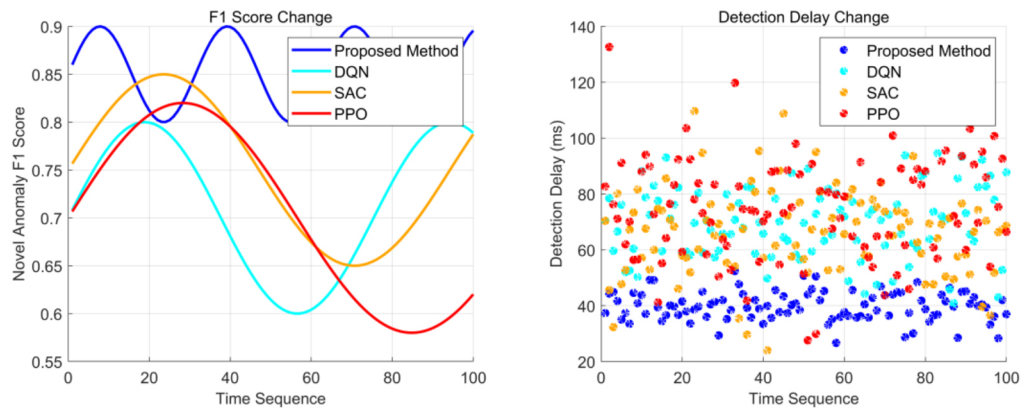


Figure 5 Comparison of model performance and real-time performance.

Table 2 Comparison of drift adaptation time and error rate of different models.

Model	Drift Adaptation Time (s)	False Positive Rate (%)	False Negative Rate (%)
Proposed Method	28	1.3	2.1
SAC	41	2.5	3.8
PPO	52	1.9	4.2
DQN	67	3.6	7.3

the proposed method has significant advantages in correction precision and collaborative work, especially in terms of correction success rate and MSE, reflecting the unique advantages of the proposed model in adaptive modeling and anti-error correction capabilities under the reinforcement learning framework. This method enables the model to effectively cope with the high uncertainty in power load data by designing a Markov decision process with random fluctuations. Multi-scale state representation and safety-constrained action space enhance the decision-making ability of the model in complex environments, ensuring that it not only has high accuracy in load anomaly identification but also avoids errors caused by overcorrection. The combination of the interpretability of the reward function and the attention-driven policy network enables the algorithm to not only have a strong learning ability when dynamically adjusting the strategy, but also to provide an interpretable decision path, which improves the transparency and reliability of the model, thereby better ensuring the improvement of correction precision and collaborative efficiency.

4.3 Evaluation of Real-Time and Generalization Ability of Anomaly Detection

To comprehensively evaluate the performance of the algorithm in the task of power load anomaly detection, especially its real-time response ability and generalization performance when processing dynamically changing load data, the experiment focuses on the relationship between the accuracy of the algorithm in identifying anomalies and the response speed, as well as its adaptability when facing load behavior pattern drift. By measuring the F1 score and detection delay dynamic changes of different algorithms on continuous time

series, and combining the model adjustment time and error rate indicators after concept drift occurs, it aims to objectively demonstrate the real-time processing efficiency and stability of each model. Figure 5 shows the core comparison results of this part of the experiment.

Figure 5 shows the dynamic relationship between the F1 score and detection delay of the four algorithms in power load anomaly detection. The peak F1 score of the proposed model reaches more than 0.89, and the detection delay is stable in the range of 20-60ms. The F1 scores of DQN, SAC, and PPO fluctuate greatly; the detection delay exceeds 80ms, and PPO even exceeds 120ms. At the 100th time point, the proposed model has an F1 score of 0.9, corresponding to a 37ms delay, while PPO has an F1 of 0.62 but requires a 66ms delay. This difference in performance stems from the essential difference in the algorithm mechanism. The Markov decision process in this paper combines multi-scale state representation to more precisely capture the time series characteristics and random fluctuations of power loads. Its safety constraint action space design effectively suppresses the delay accumulation caused by misjudgment.

Power load data often undergoes concept drift due to seasonal changes, equipment updates, or changes in operating modes, which poses a challenge to the continued effectiveness of anomaly detection models. To quantitatively evaluate the adaptability of different algorithms to such data distribution changes, this experiment simulates the load pattern drift scenario and records key performance indicators. The focus is on the time required for the model to recover to stable detection performance after drift occurs, as well as the misjudgment generated in this process. Table 2 compares the performance of each model in terms of drift adaptation time, false positive rate, and false negative rate.

Table 2 shows the performance of the four models in terms of drift adaptation time, false positive rate, and false negative

Table 3 Comprehensive comparison of interpretability verification indicators.

Evaluation Metric	Proposed Method	SAC	PPO	DQN
Attention Weight Coverage (%)	95.1	82.4	87.6	74.3
Counterfactual Explanation Consistency (%)	93.2	80.1	85.7	72.4
Constraint Trigger Explainability (%)	89.7	75.6	80.2	68.9

rate. The model in this paper takes only 28 seconds to adapt to the drift; SAC takes 41 seconds; PPO takes 52 seconds; DQN takes the longest, reaching 67 seconds. This shows that the proposed model can respond quickly to load fluctuations and abnormal behaviors and return to a stable state. In terms of false positive rate, the 1.3% of the model in this paper is also lower than that of other models, among which DQN has the highest false positive rate of 3.6%. In terms of false negative rate, the proposed model also performs well, with only 2.1%, which is lower than DQN's 7.3%, indicating that the model can more effectively avoid the omission of abnormal loads and reduce the risk of system false positives and false negatives. The proposed model can quickly adapt to changes in load data and make corrections, especially when facing rapid changes in abnormal loads, showing excellent dynamic adjustment capabilities. In contrast, models such as DQN have weaker adaptability and often take longer to adjust strategies during training, resulting in longer drift adaptation time and being prone to higher false positives and false negatives. These differences indicate that the proposed model takes into account more complex and random power system behaviors when designing and improves the overall reliability and real-time response capability of the system by applying more sophisticated state representation and reinforcement learning mechanisms.

4.4 Interpretability Verification Evaluation

The interpretability of the algorithm is crucial for power system operators to understand and trust model decisions, especially in scenarios involving safety corrections. To evaluate the interpretability component integrated in the method, three key indicators are experimentally studied: the attention weight coverage that measures the transparency of the model decision basis, the indicator that evaluates the rationality of the counterfactual explanation provided by the model, and the degree to which the safety constraints are clearly explained. Through a comprehensive comparison with other mainstream reinforcement learning algorithms on these indicators, the advantages of the verification method in improving decision transparency are verified. Table 3 summarizes the quantitative results of interpretability verification.

Table 3 shows the comprehensive comparison results of the interpretability verification indicators of the adaptive modeling and anti-error correction algorithm of abnormal behavior of power load based on reinforcement learning with the other three algorithms. For the coverage of attention weights, the proposed method performs best. In terms of counterfactual explanation consistency, the proposed method scores 93.2%, which is higher than the other three algorithms. Finally, in regard to the constraint trigger interpretability indi-

cators, the proposed method maintains a high level (89.7%), significantly higher than SAC (75.6%), PPO (80.2%), and DQN (68.9%). The data shows that the overall performance of the proposed method in terms of interpretability is good. It takes the lead in terms of interpretability due to its unique reinforcement learning framework design. In contrast, although SAC, PPO, and DQN are not weak in performance, they are insufficient in these indicators due to the lack of a design optimized for interpretability. The high coverage of attention weights enables operation and maintenance personnel to intuitively understand the key basis of model decision-making, quickly locate the root cause characteristics of abnormalities, and greatly improve the efficiency of fault diagnosis. The counterfactual explanation consistency index verifies the rationality of the explanation provided by the model, which helps operation and maintenance personnel to evaluate the potential consequences of different correction strategies, and assists them in formulating better emergency plans. The high constraint trigger interpretability ensures that the decision logic of the safety barrier module is transparent and credible, enhances the trust of operation and maintenance personnel in the automated correction system, and reduces the risk of human intervention delays caused by "black box" decision-making.

4.5 Sensitivity to Abnormality Types

The power system may encounter various types of load abnormalities, and different abnormalities have their own unique pattern characteristics (such as suddenness, persistence, slow change, etc.). To evaluate the algorithm's recognition sensitivity and correction effectiveness for different abnormal patterns, this experiment is conducted on a dataset containing five typical abnormalities (sudden load peaks, long-term load fluctuations, load drifts, sudden equipment failures, and external interference). The experiment measures the F1 score of each algorithm when detecting each type of abnormality and the proportion of successful correction (abnormal correction success rate), revealing the model's adaptability to different abnormal characteristics and processing efficiency differences, as shown in Figure 6.

Figure 6 shows the F1 scores and anomaly correction success rates of the four models under five anomaly types. The proposed model performs well under most anomaly types, especially for long-term fluctuation anomalies, where its F1 score reaches 1.0, showing extremely high anomaly recognition ability. In comparison, DQN performs relatively poorly on long-term failures, at 0.45, but has a relatively balanced performance on load drift and external interference. The F1 score of PPO is at a lower level than that of other models under most anomaly types, especially for burst

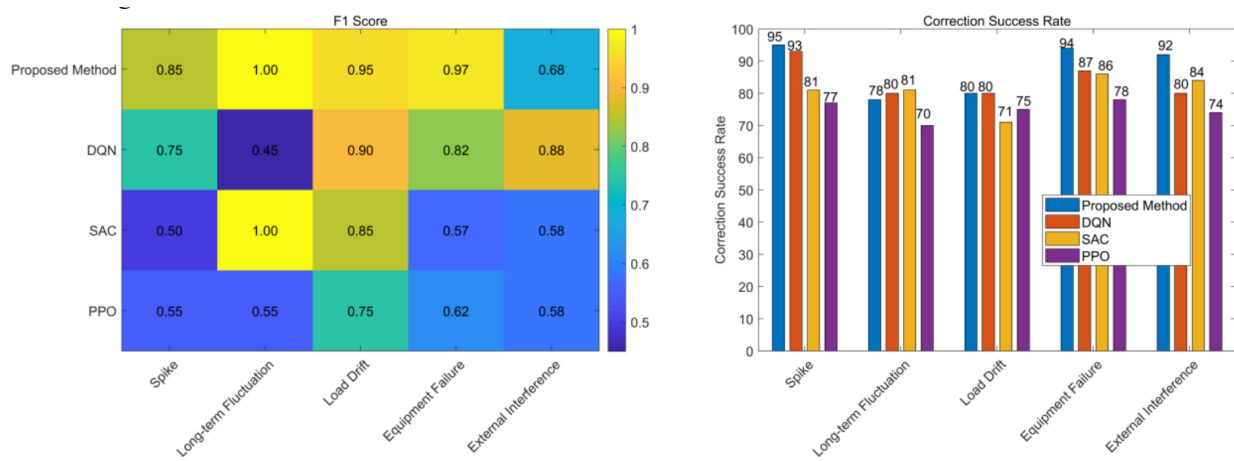


Figure 6 Performance comparison of algorithms in anomaly detection and correction.

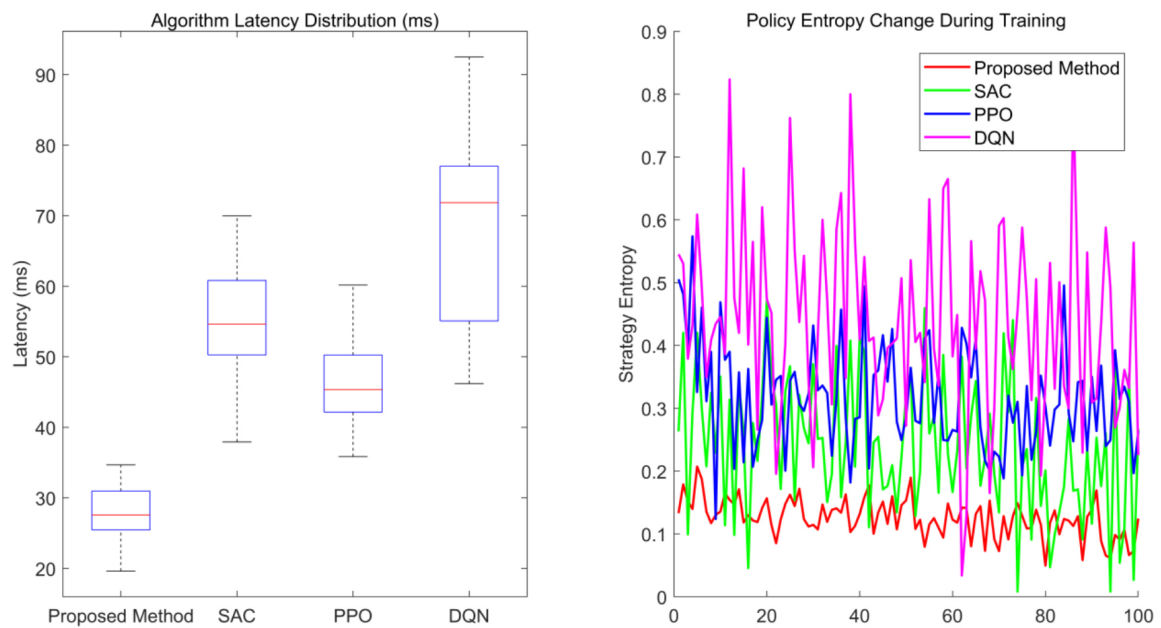


Figure 7 Delay distribution and strategy entropy change trend.

peak anomalies, where its score is only 0.55, which is poor. In terms of anomaly correction success rate, the proposed model has a success rate of 95% and 94% for burst peaks and equipment failures, respectively, which is higher than other models. The correction success rate of PPO is relatively low for multiple anomaly types, especially for long-term fluctuations and external interference, at 70% and 74%, respectively. The proposed model combines the reinforcement learning framework and the attention mechanism, which can effectively identify different anomaly types and accurately correct them, so it can maintain a high F1 score and correction success rate under most anomaly types. PPO fails to fully optimize several more complex anomaly types because the model may face greater instability during the exploration process, resulting in their anomaly correction success rate being lower than that of the model in this paper.

4.6 Dynamic Performance Evaluation of Adaptive Modeling

The ability of the algorithm to adapt to environmental changes and dynamically update strategies online is the key to its deployment in actual power systems. This experiment focuses on analyzing the dynamic performance of the model during operation, especially the response speed and decision stability of the model update strategy. The experiment studies the adaptive efficiency of the model and the stability of the strategy convergence by measuring the delay distribution required for the model to process data updates and the change trend of the entropy value of the probability distribution of the action output by the strategy network (reflecting the certainty or randomness of the strategy) over time, as shown in Figure 7.

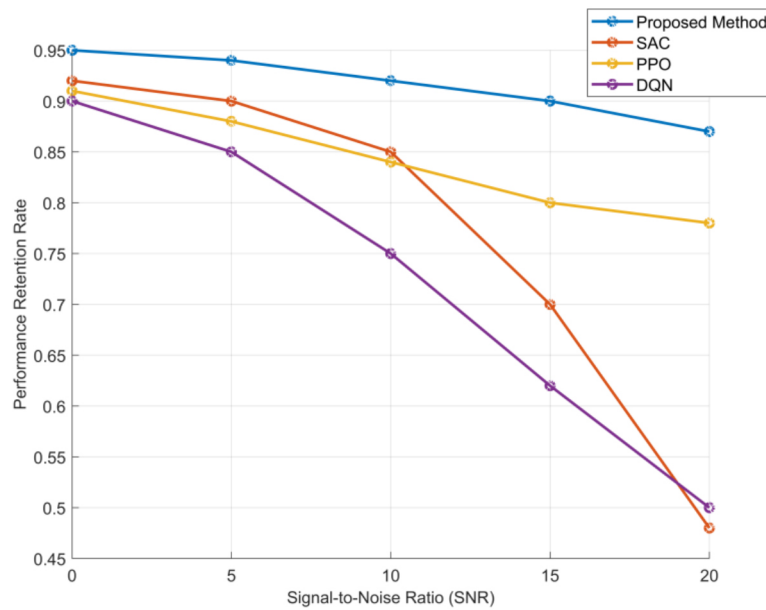


Figure 8 Model performance retention rate under different noise environments.

The delay of the proposed method is concentrated in the range of 25–32ms, which is significantly lower than that of SAC, PPO, and DQN. The delay of most samples of the proposed method is less than 31ms, while DQN has two abnormally high values of more than 85ms. The strategy entropy value of the proposed method is maintained in the range of 0.04–0.21; the entropy value of DQN fluctuates violently between 0.03–0.83, and high entropy points exceeding 0.5 frequently appear in the middle and late stages of training, indicating that the decision stability is seriously insufficient. The proposed method locks the model update within the load change rate safety threshold through dual optimization of the safety constraint action space and the attention-driven policy network, thus reducing the delay. The online policy update mechanism adopts adaptive learning rate adjustment, dynamically increases the learning rate when the load changes suddenly, and automatically reduces the learning rate during the stable period, which confirms the key role of the safety barrier module in maintaining decision stability.

4.7 Comprehensive Comparative Evaluation of Industrial Scenarios

The power load data in actual industrial environments are often accompanied by varying degrees of noise interference. To evaluate the robustness and performance retention ability of the algorithm in a noisy environment closer to the real scene, this experiment simulates noise conditions of different intensities. By gradually increasing the noise level applied to the test dataset, the attenuation of the core performance indicators of each algorithm is observed and compared. Figure 8 shows the performance retention rate change curves of the four models under different noise intensities.

In Figure 8, the horizontal axis represents different noise intensities, which are divided into five levels from 0dB to 20dB, and the vertical axis shows the performance retention rate of the four models under different noise environments. When the noise intensity is 0dB, the performance retention rate of the proposed model is the highest, reaching 0.95. As the noise intensity increases, the performance of all models decreases to varying degrees. Among them, the model proposed in this paper remains relatively stable, with the retention rate dropping to 0.94 when the noise intensity is 5dB. The further increase in noise gradually increases its impact, and it drops to 0.87 when the noise intensity is 20dB, but the performance remains at a good level, indicating that the model has strong robustness. The performance of the SAC model fluctuates greatly in a noisy environment, reflecting the high sensitivity of the model to noise. This phenomenon further shows that the robustness and performance maintenance ability of the model in this paper are closely related to the noise processing ability, model structure, and optimization strategy during training.

5. CONCLUSIONS

In view of the high uncertainty and error correction risk in power load anomaly identification, this paper proposes a reinforcement learning framework that integrates adaptive modeling and anti-error correction. A Markov decision process with random perturbations is constructed, combining multi-scale state representation to dynamically capture the law of load fluctuations; a hierarchical safety constraint action space is designed to avoid overcorrection risks, and the decision process is optimized based on an interpretable reward function. The attention-driven strategy network focuses on key features to improve decision transparency,

and the dynamic safety barrier module intercepts high-risk operations in realtime. Future technology development can take four directions: in different types of power grids, the model is verified and optimized; a lightweight model adapted to edge computing equipment is developed to achieve local real-time processing; deeply integrated with the existing energy management system and data acquisition and monitoring system, standard interface specifications are formulated; safety barrier triggering logic and manual review interaction interface are optimized to improve operation and maintenance efficiency.

REFERENCES

1. Li, S., Pandey, A., Hooi, B., Faloutsos, C., & Pileggi, L. (2021). Dynamic graph-based anomaly detection in the electrical grid. *IEEE Transactions on Power Systems*, 37(5), 3408–3422.
2. Baker, M., Fard, A. Y., Althuwaini, H., & Shadmand, M. B. (2022). Real-time AI-based anomaly detection and classification in power electronics dominated grids. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, 4(2), 549–559.
3. Bislimi, A. (2024). Comprehensive analysis of power system: exploring load factor, power balance, active load variation, and increment factors with iterative implications. *International journal of electrical and computer engineering systems*, 15(1), 105–112.
4. Patil, K. R., Karnik, S. R., & Raju, A. B. (2021). A Comparative study of DG impacts on power system stability under equal and unequal load growth scenarios. *Asian Journal for Convergence in Technology (AJCT)* ISSN-2350-1146, 7(2), 05–11.
5. Jiang, J., Li, W., Wen, Z., Bie, Y., Schwarz, H., & Zhang, C. (2021). Series arc fault detection based on random forest and deep neural network. *IEEE Sensors Journal*, 21(15), 17171–17179.
6. Pan, X., Chen, M., Zhao, T., & Low, S. H. (2022). DeepOPF: A feasibility-optimized deep neural network approach for AC optimal power flow problems. *IEEE Systems Journal*, 17(1), 673–683.
7. Chen, S. (2021). Review on supervised and unsupervised learning techniques for electrical power systems: Algorithms and applications. *IEEE transactions on electrical and electronic engineering*, 16(11), 1487–1499.
8. Wu, S., Fang, L., Zhang, J., Sriram, T. N., Coshatt, S. J., Zahiri, F., ... & Song, W. (2023). Unsupervised anomaly detection and diagnosis in power electronic networks: Informative leverage and multivariate functional clustering approaches. *IEEE Transactions on Smart Grid*, 15(2), 2214–2225.
9. Sharma, R., Joshi, A. M., Sahu, C., & Nanda, S. J. (2023). Detection of false data injection in smart grid using PCA based unsupervised learning. *Electrical Engineering*, 105(4), 2383–2396.
10. Shirzadi, N., Nizami, A., Khazen, M., & Nik-Bakht, M. (2021). Medium-term regional electricity load forecasting through machine learning and deep learning. *Designs*, 5(2), 27.
11. Dong, X., Deng, S., & Wang, D. (2022). A short-term power load forecasting method based on k-means and SVM. *Journal of Ambient Intelligence and Humanized Computing*, 13(11), 5253–5267.
12. Oprea, S. V., Bâra, A., Puican, F. C., & Radu, I. C. (2021). Anomaly detection with machine learning algorithms and big data in electricity consumption. *Sustainability*, 13(19), 10963.
13. Ahmed, A., Sajan, K. S., Srivastava, A., & Wu, Y. (2021). Anomaly detection, localization and classification using drifting synchrophasor data streams. *IEEE Transactions on Smart Grid*, 12(4), 3570–3580.
14. Takiddin, A., Ismail, M., Zafar, U., & Serpedin, E. (2022). Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids. *IEEE Systems Journal*, 16(3), 4106–4117.
15. Udo, W. S., Kwakye, J. M., Ekechukwu, D. E., & Ogundipe, O. B. (2024). Smart grid innovation: machine learning for real-time energy management and load balancing. *International Journal of Smart Grid Applications*, 22(4), 405–423.
16. Presekali, A., Ştefanov, A., Rajkumar, V. S., & Palensky, P. (2023). Attack graph model for cyber-physical power systems using hybrid deep learning. *IEEE Transactions on Smart Grid*, 14(5), 4007–4020.
17. Jahangir, H., Lakshminarayana, S., Maple, C., & Epiphanou, G. (2023). A deep-learning-based solution for securing the power grid against load altering threats by IoT-enabled devices. *IEEE Internet of Things Journal*, 10(12), 10687–10697.
18. Mohamed, A. S., & Kundur, D. (2023). On the use of reinforcement learning for attacking and defending load frequency control. *IEEE Transactions on Smart Grid*, 15(3), 3262–3277.
19. Lei, J., Gao, S., Shi, J., Wei, X., Dong, M., Wang, W., & Han, Z. (2022). A reinforcement learning approach for defending against multiscenario load redistribution attacks. *IEEE Transactions on Smart Grid*, 13(5), 3711–3722.
20. Shereen, E., Kazari, K., & Dán, G. (2023). A reinforcement learning approach to undetectable attacks against automatic generation control. *IEEE Transactions on Smart Grid*, 15(1), 959–972.
21. Gou, J., Niu, X., Chen, X., Dong, S., & Xin, J. (2025). Identification of Abnormal Electricity Consumption Behavior of Low-Voltage Users in New Power Systems Based on a Combined Method. *Energies*, 18(10), 2528.
22. Aljohani, A., AlMuhaini, M., Poor, H. V., & Binqadhi, H. M. (2024). A deep learning-based cyber intrusion detection and mitigation system for smart grids. *IEEE Transactions on Artificial Intelligence*, 5(8), 3902–3914.
23. Cooper, A., Bretas, A., & Meyn, S. (2023). Anomaly detection in power system state estimation: Review and new directions. *Energies*, 16(18), 6678.
24. Alam, K., Imran, M. A., Mahmud, U., & Fathah, A. A. (2024). Cyber Attacks Detection And Mitigation Using Machine Learning In Smart Grid Systems. *Journal of Science and Engineering Research*, 1(01), 38–55.
25. Das, P., Shuvro, R. A., Povinelli, K., Sorrentino, F., & Hayat, M. M. (2022). Mitigating cascading failures in power grids via Markov decision-based load-shedding with DC power flow model. *IEEE Systems Journal*, 16(3), 4048–4059.
26. Qin, Z., Liu, D., Hua, H., & Cao, J. (2021). Privacy preserving load control of residential microgrid via deep reinforcement learning. *IEEE Transactions on Smart Grid*, 12(5), 4079–4089.
27. Zhao, N., & You, F. (2021). Toward carbon-neutral electric power systems in the new york state: A novel multi-scale bottom-up optimization framework coupled with machine learning for capacity planning at hourly resolution. *ACS Sustainable Chemistry & Engineering*, 10(5), 1805–1821.
28. Meng, S., Li, C., Peng, W., & Tian, C. (2023). Empirical mode decomposition-based multi-scale spectral graph convolution network for abnormal electricity consumption detection. *Neural Computing and Applications*, 35(13), 9865–9881.
29. Ma, Y., Oslebo, D., Maqsood, A., & Corzine, K. (2021). Pulsed-power load monitoring for an all-electric ship: Utilizing the fourier transform data-driven deep learning approach. *IEEE Electrification Magazine*, 9(1), 25–35.

30. Peng, S., Zhang, Z., Deng, R., & Cheng, P. (2023). Localizing false data injection attacks in smart grid: A spectrum-based neural network approach. *IEEE Transactions on Smart Grid*, 14(6), 4827–4838.